Osteoarthritis and Cartilage 19 (2011) 483–492

Osteoarthritis and Cartilage

Outcome measures in placebo-controlled trials of osteoarthritis: responsiveness to treatment effects in the REPORT database

R.H. Dworkin †*, S. Peirce-Sandner †, D.C. Turk ‡, M.P. McDermott §, A. Gibofsky ||, L.S. Simon ¶, J.T. Farrar #, N.P. Katz ††‡‡

† Department of Anesthesiology, University of Rochester School of Medicine and Dentistry, Rochester, NY, USA
‡ Department of Anesthesiology, University of Washington, Seattle, WA, USA
§ Department of Biostatistics and Computational Biology, University of Rochester School of Medicine and Dentistry, Rochester, NY, USA
|| Division of Rheumatology, Hospital for Special Surgery, Weill Medical College of Cornell University, New York, NY, USA
¶ SDG, LLC, Cambridge, MA, USA
# Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania, PA, USA
†† Analgesic Solutions, Natick, MA, USA
‡‡ Tufts University, Boston, MA, USA

ARTICLE INFO

Article history:
Received 24 December 2010
Accepted 14 February 2011

Keywords:
Outcome measures
Randomized clinical trials
Osteoarthritis
Responsiveness
Treatment effect
Standardized effect size

SUMMARY

Introduction: Treatment response in randomized clinical trials (RCT) of osteoarthritis (OA) has been assessed by multiple primary and secondary outcomes, including pain, function, patient and clinician global measures of status and response to treatment, and various composite and responder measures. Identifying outcome measures with greater responsiveness to treatment is important to increase the assay sensitivity of RCTs.
Objective: To assess and compare the responsiveness of different outcome measures used in placebo-controlled RCTs of OA.
Search strategy: The Resource for Evaluating Procedures and Outcomes of Randomized Trials database includes placebo-controlled clinical trials of pharmacologic treatments (oral, topical, or transdermal) for OA identified from a systematic literature search of RCTs published or publicly available before August 5, 2009, which was conducted using PubMed, the Cochrane collaboration, publicly-available websites, and reference lists of retrieved publications.
Data collection and analysis: Data collected included: (1) pain assessed with single-item ratings and the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain subscale; (2) patient and clinician global measures of status, improvement, and treatment response; (3) function assessed by the WOMAC function subscale; (4) stiffness assessed by the WOMAC stiffness subscale; and (5) the WOMAC and Lequesne Algofunctional Index composite outcomes. Measures were grouped according to the total number of response categories (i.e., $<$10 categories or $\geq$10 categories). The treatment effect (difference in mean change from baseline between the placebo and active therapy arms) and standardized effect size (SES) were estimated for each measure in a meta-analysis using a random effects model.
Results: There were 125 RCTs with data to compute the treatment effect for at least one measure; the majority evaluated non-steroidal anti-inflammatory drugs (NSAIDs), followed by opioids, glucosamine and/or chondroitin, and acetaminophen. In general, the patient-reported pain outcome measures had comparable responsiveness to treatment as shown by the estimates of treatment effects and SES. Treatment effects and SESs were generally higher for patient-reported global measures compared with clinician-rated global measures but generally similar for the WOMAC and Lequesne composite measures.
Conclusions: Comparing different outcome measures using meta-analysis and selecting those that have the greatest ability to identify efficacious treatments may increase the efficiency of clinical trials of treatments for OA. Improvements in the quality of the reporting of clinical trial results are needed to facilitate meta-analyses to evaluate the responsiveness of outcome measures and to also address other issues related to assay sensitivity.

* Address correspondence and reprint requests to: Robert H. Dworkin, Department of Anesthesiology, University of Rochester Medical Center, 601 Elmwood Ave., Box 604, Rochester, NY 14642, USA. Tel: 1-585-275-8214; Fax: 1-585-276-0122.
E-mail address: robert_dworkin@urmc.rochester.edu (R.H. Dworkin).

1063-4584/$ – see front matter
doi:10.1016/j.joca.2011.02.020

Introduction

An appreciable percentage of patients with osteoarthritis (OA) are refractory to existing analgesic treatments, and the patients who do respond to these treatments often obtain only partial relief of their pain[1,2]. Considerable effort is therefore being devoted to the development of new treatments for OA and to conducting randomized clinical trials (RCTs) to evaluate their efficacy and safety. In designing these trials, it is critically important that methodological factors are identified that might improve their assay sensitivity, which has been defined as "the ability to distinguish an effective treatment from a less effective or ineffective treatment"[3]. Assuming that the treatment studied is efficacious, RCTs with greater assay sensitivity are less likely to have falsely negative study results and can detect treatment effects with smaller sample sizes. This not only hastens the time to study completion and reduces costs, but also exposes fewer subjects to the unknown risks of novel treatments.

An essential aspect of the assay sensitivity of a clinical trial involves the responsiveness to treatment of its outcome measures. The importance of assay sensitivity in identifying efficacious treatments as efficiently as possible provides a compelling rationale for identifying and then selecting measures that have the greatest responsiveness (assuming other characteristics of the measures do not offset this, for example, lack of clinical importance or substantially increased patient burden).

Response to treatment in RCTs of patients with OA has been measured by patient-reported assessments of pain, function, and stiffness as well as by patient and clinician global evaluations of disease status and response to treatment[4–6]. Various visual analog scales (VAS) and numeric rating scales (NRS) as well as disease-specific outcome measures, such as the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)[7] and the Lequesne Algofunctional Index[8], have been widely used as primary and secondary outcome measures in OA trials. To encourage standardization among the diverse measures that are available, the Osteoarthritis Research Society International (OARSI) and Outcome Measures in Rheumatology (OMERACT) have recommended core outcome domains and measures for OA clinical trials[9–12]. For RCTs of chronic pain conditions in general, the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) has recommended consideration of specific outcome domains[13], measures[14], approaches to developing new measures[15], and strategies for evaluating the clinical importance of treatment outcomes[16,17].

IMMPACT has also created the Resource for Evaluating Procedures and Outcomes of Randomized Trials (REPORT) to examine relationships between clinical trial research methods and study outcomes and thereby contribute to the development of an evidence-based approach to analgesic clinical trial design[18,19]. REPORT consists of comprehensive databases of RCTs of acute and chronic pain conditions (e.g., neuropathic pain, fibromyalgia, chronic low back pain, and acute post-operative pain), which are at present limited to trials of pharmacologic treatments. In this article, we evaluate and compare the responsiveness to treatment of commonly used outcome measures in placebo-controlled RCTs of pharmacologic treatments in the REPORT database of OA trials.

Materials and methods

Literature search

The REPORT database of OA clinical trials includes placebo-controlled trials of pharmacologic treatments identified from a systematic literature search of RCTs published or publicly available before August 5, 2009, which was conducted using PubMed, Cochrane collaboration systematic reviews, publicly-available websites, and references from published reports of trials that met inclusion criteria, and other retrieved publications[20]. The search terms included "osteoarthritis", "degenerative joint disease", "coxarthrosis", and "gonarthrosis," with limits of "Randomized Controlled Trial", "Human", and "English" applied. Only trials that met the following criteria were included: (1) results reported in publicly-available sources, including publications and websites (e.g., www.clinicaltrialresults.org); (2) evaluated oral, topical, or transdermal pharmacologic treatments; (3) had treatment durations of at least 7 days; (4) used a parallel group design; (5) included patients with OA of the knee or hip; and (6) were placebo-controlled and double blinded (except for one single blind trial). Clinical trials reported only in abstract form were not included.

Information was extracted on standard forms and entered into a spreadsheet. Variables collected from each trial that were used in the present analyses included the following, when available: (1) eligibility criteria, including joint(s) studied; (2) active treatments; (3) baseline and endpoint mean values and either the respective standard deviation (SD) or information from which this SD could be derived (e.g., standard error, confidence interval); (4) specific outcome measures and scales used for all primary and secondary endpoints related to pain, function, stiffness, and patient and clinician global assessments of status, improvement, and treatment response; and (5) responder outcomes based on pain reduction, patient and clinician global or treatment response assessments, and OMERACT–OARSI responder criteria[5,11]. The statistical significance for the comparison of each active treatment group with the placebo group was recorded for all primary and secondary endpoints.

Study outcomes

Data from primary and secondary endpoints using 0–10 NRS, 0–10 cm VAS, and other measures were transformed to a 0–100 scale[21]. Although NRSs and VASs may have somewhat different psychometric properties, responses to these two types of measures of pain intensity are highly correlated and there is no evidence that their responsiveness to change and to treatment effects differs[14].

The placebo group and treatment group mean changes from baseline and standardized effect sizes (SESs) were determined as follows:

1. Placebo group mean change from baseline, as reported or computed as the difference in mean responses at the baseline and final visits.
2. Active treatment group mean change from baseline, as reported or computed as the difference in mean responses at the baseline and final visits.
3. Treatment effect: active treatment group mean change from baseline − placebo group mean change from baseline.
4. Pooled SD:

$$\frac{(N_{Treatment}-1)\times(SD_{Treatment})^2+(N_{Placebo}-1)\times(SD_{Placebo})^2}{N_{Treatment}+N_{Placebo}-2}$$

5. SES: treatment effect/pooled SD

When only final visit means were available (i.e., no mean baseline or change from baseline values were present), expressed as either actual mean values or mean values adjusted for baseline, the treatment effect was calculated by subtracting the placebo

group final mean value from the active treatment group final mean value. The SES was calculated only for trials that reported the corresponding measure of variability (e.g., SD) of the change from baseline[22] or of the final values adjusted for baseline.

Analyses focused on RCTs of efficacious treatments, which were considered those medications that are "recommended treatments" for OA in prominent national and international guidelines[23,24]. Given the difficulty of establishing which other treatments truly lack efficacy, supplementary analyses are presented in Appendix I for RCTs examining all treatments irrespective of whether they are recommended or not. For trials with multiple arms examining the same treatment (e.g., different dosages or titration schedules), a mean value for each treatment was calculated by computing the average response, weighted according to sample size, across all arms of the same treatment within a trial so that each active treatment contributed only one value per trial for each outcome measure.

The outcome measures were included in the analyses of treatment effects and SESs if there were at least five RCTs with sufficient data[25]. Pain-related outcome measures included: (1) "pain", representing spontaneous pain, pain at rest, or otherwise undesignated pain; (2) "active pain", representing pain during a weight-bearing activity, for example, when walking or standing, but not in response to passive movement by an examiner or clinician; (3) the WOMAC pain subscale[7]; and (4) item one of the WOMAC pain subscale (pain walking on a flat surface)[7]. Patient and clinician measures were based on global assessments (e.g., of OA status or overall improvement) and on response to treatment of OA. Function and stiffness endpoints were measured by the WOMAC function and stiffness subscales[7], and composite outcomes were represented by the WOMAC total score[7] and the Lequesne Algofunctional Index[8,26], a 10-item composite measure of pain, stiffness, walking distance, and other aspects of function. All of these measures were also evaluated according to the number of response categories for ratings made by the patient or clinician during the trial, specifically (1) scales with 10 response categories or more (e.g., NRS with a 0–10 scale or VAS with a 0–100 mm scale); and (2) scales with fewer than 10 response categories (e.g., Likert scale with a 0–3 verbal rating scale). For measures that consist of multiple individual rating scales each having fewer than 10 response categories, the number of response categories was considered the total number of possible categories; for example, a measure composed of four 0–3 Likert scales with a maximum total score of 12 was considered a scale with $\geq$10 response categories. Responder outcomes were classified according to the instrument used to categorize responders and non-responders, specifically, pain (i.e., $\geq$30% reduction), patient and clinician global ratings and treatment response assessments, and OMERACT–OARSI responder criteria[5,11], which are based on absolute and percentage improvements in pain, function and patient global assessments.

Statistical analyses

Due to the expected heterogeneity among the different studies (including eligibility criteria, intervention studied, evaluation protocol, concomitant treatments, and other factors), overall estimates of treatment effects and SESs for each outcome measure were obtained from random effects models that treated study as a random effect[27,28]. These overall estimates were computed as weighted averages of the individual study estimates, with the weights being inversely proportional to the estimated variances of the individual study estimates[27,28]. Responder measures with binary outcomes (i.e., responder vs non-responder, however defined) were evaluated in the same manner, with the overall percentage of responders estimated by the weighted average of the study-specific estimated percentages, with the weights being inversely proportional to the estimated variances of the individual study estimates[27,28].

The number of response categories used by outcome measures could be associated with differences in either treatment effects or SESs, for example, fewer response categories could be associated with less responsiveness to change. To evaluate this possibility, mixed effects models with number of scale categories as a fixed effect (0 = scales with <10 response categories, 1 = scales with $\geq$10 response categories) and study as a random effect was used. Because the effect of number of scale response categories was significant for clinician global status for treatment effect and SES and showed trends for the other measures that used both of these two groups of number of response categories, treatment effects and SESs were summarized overall as well as according to number of scale categories when an outcome measure had adequate data for both groups of number of response categories.

The estimates of treatment effect and SES for selected measures were compared using a mixed effects model with study as a random effect and the measure of interest as a fixed effect. The measures selected for comparison were the most commonly used outcome measures that were similar with respect to the underlying construct being assessed (i.e., pain, global status, overall composite outcome): (1) spontaneous pain vs WOMAC pain; (2) patient vs clinician global assessment; and (3) WOMAC total score vs Lequesne index. In the comparisons of these selected measures, when there were studies that had values for both of the measures, the measure with the greatest number of values in each of the analyses was discarded from the study to preserve independence. For the random effects models, when a trial included two or more different active treatments, only one active treatment arm per placebo group was retained in order to preserve independence. A single active treatment arm was randomly selected from trials with two or more different active treatments and any other active treatment arms were discarded.

For each outcome measure, the following data are presented: (1) estimates from the random effects models overall and for each of the two groups of measures differing in number of response categories, when data from five or more trials were available[25]; (2) overall unweighted means across all arms to show treatment effects and SESs for the greatest number of treatment arms, including those for which no measure of variability was available. The numbers of treatment arms available for the random effects models are lower than the number of arms used in calculating the unweighted means because only one active arm per trial was used and the random effects model requires an estimate of variance that was not always available. The sample sizes for estimates of the SESs are often lower than for the treatment effects because SES was computed only for studies that reported variance adjusted for baseline values[22]. In comparing selected measures (e.g., spontaneous pain and WOMAC pain) for treatment effects and for SESs, the sample size is further reduced because one measure was dropped from studies that reported both measures in order to preserve independence. In the following presentation and discussion of the results, we focus on estimates from the random effects models.

Results

A total of 1774 articles were retrieved from the PubMed search, Cochrane collaboration reviews related to pharmacotherapy for OA, and the clinicaltrialresults.org website, from which 167 blinded, randomized, and placebo-controlled trials of oral, topical, and transdermal therapies for OA of the knee and/or hip were

Table I
Active treatment groups according to outcome measure and treatment type for trials with a non-missing value for treatment vs placebo difference or responder outcome

Measure	Total number of active treatment groups	Non-steroidal anti-inflammatory drug, N (%)*	Acetaminophen (paracetamol), N (%)	Glucosamine/chondroitin†, N (%)	Opioid analgesic‡, N (%)	Other, N (%)
Pain	83	45 (54.2)	0 (0)	11 (13.3)	11 (13.3)	16 (19.3)
Pain (with activity)	30	24 (80.0)	1 (3.3)	2 (6.7)	1 (3.3)	2 (6.7)
WOMAC pain subscale	99	63 (63.6)	3 (3.0)	4 (4.0)	11 (11.1)	18 (18.2)
WOMAC pain walking§	21	19 (90.5)	0 (0)	0 (0)	2 (9.5)	0 (0)
Patient global rating	66	53 (80.3)	1 (1.5)	3 (4.5)	4 (6.1)	5 (7.6)
Patient response to therapy	20	17 (85.0)	0 (0)	2 (10.0)	1 (5.0)	0 (0)
Clinician global rating	50	44 (88.0)	0 (0)	2 (4.0)	2 (4.0)	2 (4.0)
Clinician response to therapy	11	10 (90.9)	0 (0)	0 (0)	1 (9.1)	0 (0)
Lequesne algofunctional index	32	16 (50.0)	1 (3.1)	7 (21.9)	0 (0)	8 (25.0)
WOMAC total score	65	36 (55.4)	3 (4.6)	7 (10.8)	5 (7.7)	14 (21.5)
WOMAC function subscale	93	59 (63.4)	4 (4.3)	8 (8.6)	8 (8.6)	14 (15.1)
WOMAC stiffness subscale	82	51 (62.2)	2 (2.4)	7 (8.5)	8 (9.8)	14 (17.1)
≥30 % pain reduction responder	14	4 (28.6)	1 (7.1)	2 (14.3)	4 (28.6)	3 (21.4)
Patient responder	61	44 (72.1)	1 (1.6)	7 (11.5)	4 (6.6)	5 (8.2)
Clinician responder	35	26 (74.3)	0 (0)	5 (14.3)	3 (8.6)	1 (2.9)
OMERACT–OARSI responder	18	12 (66.7)	2 (11.1)	4 (22.2)	0 (0)	0 (0)

* Values for numbers of active treatment groups and percentages, which may not sum to 100% due to rounding.
† Individually or in combination.
‡ Includes tramadol or tramadol in combination with acetaminophen.
§ Item one of the WOMAC pain subscale.

identified. Of these 167 reports, 125 studies representing 184 different active treatment arms had sufficient data to compute the treatment effect for at least one measure. The majority of these RCTs examined only knee OA (66%), followed by the combination of knee OA and hip OA (30%), and hip OA only (5%). The categories of recommended treatments examined in these trials were predominantly non-steroidal anti-inflammatory drugs (NSAIDs) followed by opioid analgesics, glucosamine and/or chondroitin, and acetaminophen/paracetamol (Table I).

Pain

For the pain-related outcome measures presented in Table II, estimated treatment effects ranged from 7.6 to 9.5 (all outcomes converted to a 0–100 scale) and estimated SESs ranged from 0.21 to 0.45. Treatment effect estimates were generally comparable between typically single-item NRS or VAS pain intensity measures and the WOMAC five-item pain subscale. The SES for the WOMAC pain subscale (0.45) was appreciably higher than the SES for the measures of spontaneous pain (0.27), but this difference was not statistically significant ($p = 0.09$; Table V) and was based on data from 32 studies (13 for spontaneous pain, 19 for the WOMAC pain scale).

Responder analyses based on a reduction in pain of ≥30% from baseline to endpoint showed a treatment effect of 16.1 (i.e., the difference in percentages of responders between the active treatment and placebo arms) for scales with ≥10 categories (there were inadequate data for scales with <10 categories). These responder analyses were based on a variety of different pain outcome measures, including spontaneous pain, pain with activity, WOMAC pain subscale scores, and pain walking (item one of the WOMAC pain subscale).

Patient and clinician global measures and response to treatment measures

Most trials in the database used patient and clinician global measures of disease activity to assess response, typically as secondary endpoints. The range in estimates for patient and clinician global measures was 5.6–13.2 for treatment effects and 0.20–0.68 for SESs (Table III). The estimates of treatment effect and SES were significantly higher for patient global measures compared with clinician global measures using scales with ≥10 categories, but the differences were not significant for scales with <10 categories (Table V). For patient measures of response to treatment, five trials provided the basis for estimates of treatment

Table II
Treatment vs placebo group differences and SESs for pain-related outcome measures for recommended treatments

Measure	Pain			Pain (with activity)			WOMAC pain subscale			WOMAC pain walking			≥30% pain reduction (% patients)		
	N*	Mean	95% CI†	N	Mean	95% CI	N	Mean	95% CI	N	Mean	95% CI	N	Mean	95% CI
Treatment vs placebo group difference															
All scales, random effects model‡	29	9.4	7.2, 11.6	11	7.7	3.6, 11.8	36	8.4	6.3, 10.5	6	9.5	6.7, 12.3	6	16.1	4.4, 27.8
Scales ≥10 categories, random effects model‡	27	8.7	6.9, 10.5	10	7.6	3.1, 12.1	36	8.4	6.3, 10.5				6	16.1	4.4, 27.8
All scales, unweighted, all arms§	67	11.4		28	10.3		81	8.9		21	9.9		11	15.7	
SES															
All scales, random effects model‡	13	0.27	0.11, 0.43	7	0.22	0.08, 0.36	26	0.40	0.27, 0.53	6	0.39	0.25, 0.52			
Scales ≥10 categories, random effects model‡	13	0.27	0.11, 0.43	6	0.21	0.05, 0.37	26	0.40	0.27, 0.53						
All scales, unweighted, all arms§	19	0.29		8	0.25		44	0.43		9	0.45				

* Number of active treatment arms (one value per treatment).
† CI = confidence interval.
‡ Estimate from random effects model with one randomly selected treatment included per trial.
§ Confidence intervals are not provided because their calculation would assume the statistical independence of the results from all of the trials, which does not hold in this case due to the inclusion of multiple treatment comparisons with the same placebo group in some of the trials.

Table III
Treatment vs placebo group differences and SESs for patient- and clinician-rated outcome measures for recommended treatments

Measure	Patient global			Patient treatment response			Patient responder (% patients)			Clinician global			Clinician treatment response			Clinician responder (% patients)		
	N*	Mean	95% CI†	N	Mean	95% CI	N	Mean	95% CI	N	Mean	95% CI	N	Mean	95% CI	N	Mean	95% CI
Treatment vs placebo group difference																		
All scales, random effects model‡	26	10.7	8.5, 12.9	5	17.1	13.2, 21.0	41	23.1	19.8, 26.4	18	8.6	6.6, 10.6				28	24.5	19.6, 29.4
Scales ≥10 categories, random effects model‡	20	10.1	7.5, 12.7							7	5.6	3.4, 7.8						
Scales < 10 categories, random effects model‡	6	13.2	10.0, 16.4	5	17.1	13.2, 21.0	41	23.1	19.8, 26.4	11	10.9	8.8, 13.0				28	24.5	19.6, 29.4
All scales, unweighted, all arms§	61	12.0		20	15.5		56	23.1		48	10.9		11	18.4		34	24.5	
SES																		
All scales, random effects model‡	19	0.38	0.27, 0.49	5	0.68	0.48, 0.88				15	0.34	0.24, 0.45						
Scales ≥10 categories, random effects model‡	15	0.38	0.24, 0.52							6	0.20	0.14, 0.26						
Scales < 10 categories, random effects model‡				5	0.68	0.48, 0.88				9	0.46	0.33, 0.59						
All scales, unweighted, all arms§	44	0.44		9	0.66					33	0.40		7	0.74				

* Number of active treatment arms (one value per treatment).
† CI = confidence interval.
‡ Estimate from random effects model with one randomly selected treatment included per trial.
§ Confidence intervals are not provided because their calculation would assume the statistical independence of the results from all of the trials, which does not hold in this case due to the inclusion of multiple treatment comparisons with the same placebo group in some of the trials.

effect (17.1) and SES (0.68) for scales with <10 response categories. There were inadequate data to compute these estimates for scales with ≥10 categories and the clinician measures of response to treatment.

All of the patient and clinician responder outcomes were based on scales with <10 categories. Treatment effects (i.e., the difference in percentages of responders between the active treatment and placebo arms) for patient (23.1) and clinician (24.5) responder outcomes were very similar.

Function and composite measures

For the function, stiffness, and composite outcome measures, the estimates of treatment effects (5.3–8.3) and SESs (0.25–0.37) were relatively modest (Table IV), and the SESs for the Lequesne index and WOMAC total score were similar (Table V). There were 11 studies with a total of 18 different active treatment arms that provided data for the OMERACT–OARSI responder criteria. The estimated treatment effect (i.e., the difference in percentages of responders between the active treatment and placebo arms) was 12.7, which was somewhat lower than the treatment effects for the responder analyses of ≥30% pain reduction presented in Table II and the responder outcomes based on patient and clinician measures presented in Table III.

Discussion

We conducted a meta-analysis of the responsiveness of the outcome measures that are used most frequently in RCTs of OA. These measures include patient-reported assessments of pain, physical function, stiffness, global status, treatment response, and

Table IV
Treatment vs placebo group differences and SESs for function, stiffness, and composite outcome measures for recommended treatments

Measure	WOMAC function subscale			Lequesne algofunctional index			WOMAC stiffness subscale			WOMAC total score			OMERACT–OARSI responder (% patients)		
	N*	Mean	95% CI†	N	Mean	95% CI	N	Mean	95% CI	N	Mean	95% CI	N	Mean	95% CI
Treatment vs placebo group difference															
All scales, random effects model‡	36	6.8	5.5, 8.2	14	5.3	3.5, 7.1	33	7.4	5.7, 9.2	24	5.5	4.4, 6.6	11	12.7	7.8, 17.6
Scales ≥10 categories, random effects model‡	36	6.8	5.5, 8.2	14	5.3	3.5, 7.1	22	8.3	5.9, 10.7	24	5.5	4.4, 6.6			
Scales < 10 categories, random effects model‡							11	5.9	4.1, 7.8						
All scales, unweighted, all arms§	79	7.6		24	5.7		68	8.2		55	6.9		18	13.2	
SES															
All scales, random effects model‡	26	0.33	0.25, 0.41	6	0.34	0.20, 0.48	25	0.32	0.23, 0.41	15	0.30	0.24, 0.36			
Scales ≥10 categories, random effects model‡	26	0.33	0.25, 0.41	6	0.34	0.20, 0.48	16	0.37	0.25, 0.49	15	0.30	0.24, 0.36			
Scales < 10 categories, random effects model‡							9	0.25	0.17, 0.33						
All scales, unweighted, all arms§	44	0.35		8	0.37		43	0.34		27	0.34				

* Number of active treatment arms (one value per treatment).
† CI = confidence interval.
‡ Estimate from random effects model with one randomly selected treatment included per trial.
§ Confidence intervals are not provided because their calculation would assume the statistical independence of the results from all of the trials, which does not hold in this case due to the inclusion of multiple treatment comparisons with the same placebo group in some of the trials.

Table V
Comparison of pain, global, and composite outcome measures for recommended treatments

Measure	Pain		WOMAC pain subscale				Patient global		Clinician global				WOMAC total score		Lequesne			
	N*	Mean	N	Mean	Difference (95% CI†)	P‡	N	Mean	N	Mean	Difference (95% CI)	P	N	Mean	N	Mean	Difference (95% CI)	P
Treatment vs placebo group difference																		
All scales, random effects model§	29	9.6	26	9.4	0.2 (−3.2, 3.6)	.90	11	11.8	18	8.7	3.1 (−0.5, 6.7)	.09	22	5.7	14	5.2	0.6 (−1.5, 2.6)	.40
Scales ≥10 categories, random effects model§	27	9.0	27	9.4	0.3 (−2.9, 3.6)	.83	13	12.3	7	5.8	6.5 (2.0, 11.1)	.01	22	5.7	14	5.2	0.6 (−1.5, 2.6)	.40
Scales <10 categories, random effects model§							6	13.2	8	10.5	2.7 (−1.3, 6.7)	.19						
SES																		
All scales, random effects model§	13	0.27	19	0.45	0.19 (−0.03, 0.40)	.09	7	0.42	15	0.34	0.07 (−0.12, 0.26)	.55	14	0.31	6	0.34	0.03 (−0.11, 0.17)	.65
Scales ≥10 categories, random effects model§	13	0.27	19	0.45	0.19 (−0.03, 0.40)	.09	9	0.50	6	0.20	0.30 (0.07, 0.43)	.01	14	0.31	6	0.34	0.03 (−0.11, 0.17)	.65
Scales <10 categories, random effects model§							4	0.41	8	0.47	.06 (−0.15, 0.27)	.56						

* Number of active treatment arms (one value per treatment).
† CI = confidence interval.
‡ *P* value for difference between measures.
§ All values estimated from random effects model with one randomly selected treatment included per trial. When both measures were included in the same trial, the measure with the highest N in each of the three analyses (all scales, scales with ≥10 categories, scales with <10 categories) was deleted from that trial.

composite outcome, clinician-rated global assessments, and various responder outcomes. For patient-reported pain outcomes, there were generally comparable treatment effect estimates for the different measures, including single VAS or NRS ratings of overall pain or pain with activity (e.g., pain walking) and the multi-item WOMAC pain subscale. However, the mean SES for pain rated on single scales with ≥10 response categories was considerably lower than the mean SES for the WOMAC pain subscale (0.27 vs 0.45; $p = .09$, 95% confidence interval (CI) for difference = −0.03, 0.40). Although this difference was not statistically significant and was based on a total of only 32 trials, SES differences of this magnitude would have important implications with respect to sample size, with considerably fewer patients being required for adequate statistical power if the WOMAC pain subscale were to be used rather than a single pain rating. However, these results need to be interpreted with caution because it is possible that confounding factors — that is, systematic differences between trials using a single pain rating and those using the WOMAC pain subscale — may have influenced these results.

The conclusion that different patient-reported outcome measures of pain severity may have generally comparable responsiveness to treatment is consistent with the results of data analyses from single clinical trials[29,30]; for example, the difference between improvement in pain on the WOMAC pain subscale and on a VAS following knee lavage was not significant[29]. In research comparing VAS and Likert versions of the WOMAC[31,32], and in other studies of OA pain assessed using VAS and Likert scales[33,34], generally comparable responsiveness to change of these different rating scales has been found. Different pain measures and scales may not always be interchangeable[35], however, and there are circumstances in which one type of assessment might be preferred[14]. Although we found non-significant differences in favor of the WOMAC pain subscale, which will need to be examined in future research, considered together with the results of previous research, our analyses suggest that different measures of pain in patients with OA may have generally comparable ability to identify efficacious treatments.

The analyses of global outcome measures showed larger treatment effects and a trend toward larger SESs for patient-reported vs clinician-rated measures (but only for measures with ≥10 response categories). This result may not be surprising. It is likely that global assessments made by clinicians are based, in large part, on what patients report to them, and improvement in pain, which is a subjective experience, appears to account for a major portion of the variation in patient global assessments of outcome and treatment satisfaction[36]. These considerations[5,13,14] and our data suggest that patient global measures are likely to provide more valid and responsive outcomes in analgesic RCTs than clinician global measures in most circumstances.

Treatment effects and SESs were generally lower or comparable for the function and composite measures compared with the pain, global, and responder outcomes, which is consistent with the results of other studies in patients with OA[29,34,37]. Treatment effects for the OMERACT–OARSI responder criteria were somewhat lower than for the other responder outcomes, which were typically based on single-item pain or global ratings. Multidimensional measures of outcome — such as the WOMAC, Lequesne index, and OMERACT–OARSI responder criteria — may provide a more comprehensive assessment of the patient's overall response to treatment, but it is not clear why this might lead to lower responsiveness to treatment effects than found with unidimensional measures. One possibility is that existing medications for pain in OA have analgesic effects that are generally modest and do not reduce pain to low enough levels for improvements in function to become apparent. In addition, some of these treatments may not have meaningful benefits on the additional outcome dimensions included in the composite outcome measures. Of course, these observed differences also may be due to chance or confounding.

Our results are based on a meta-analysis of clinical trials that were conducted using different research designs, treatments, and outcomes, an approach that has also been used recently to evaluate the "discriminating power" of outcome measures in clinical trials of fibromyalgia[38]. A different approach to examining the assay sensitivity of outcome measures involves evaluating treatment effects and SESs in a single clinical trial in which each patient completes all of the measures and patient-level data, rather than the group means used in our analyses, provide the basis for comparing measures[29,39–41]. However, the generalizability of such results is potentially limited by specific features of the clinical trial,

including patient demographic and clinical characteristics and study methodology (e.g., trial duration[42]) as well as the specific treatment examined. The present meta-analysis provides a comparative evaluation of the responsiveness of outcome measures across a broad range of patients, clinical trial characteristics, and treatments.

It is important to emphasize that the heterogeneity of the sample of trials we examined is also a limitation of our analyses. The treatment effects and SESs for the different outcome measures could reflect not only potential differences among the measures in responsiveness to treatment but also differences among trials in methods, treatment efficacy and safety, imputation of missing data, study duration, and random variation[19,43,44]. For example, we did not adjust for whether trials used a flare design, which has recently been shown to accentuate the treatment effects of NSAIDs[21], perhaps by enriching for those patients who are most likely to respond to treatment; differences in the use of this design across the outcome measures we examined could have influenced our results. In addition, the frequency with which the outcome measures we examined were administered varied greatly among trials; for example, some RCTs conducted pain ratings on a daily basis and examined the means of such multiple ratings, whereas others only captured pain weekly or monthly and examined single ratings. Measures that use multiple assessments — on different occasions or within a single measure, for example, the five items of the WOMAC pain subscale — generally have greater reliability, which might be associated with increased responsiveness to treatment effects.

There are other important limitations of our analyses. We considered glucosamine and chondroitin to be efficacious treatments, although recent evidence suggests that they might not be[45] (this conclusion, however, has been disputed[46]). In addition, it is widely recognized that negative trials are less likely to be published, and our analyses were limited to published and publicly-available RCTs; estimates of treatment effects based on the published literature are therefore likely to be higher than if all RCTs could be examined[47,48] and this could have influenced our results. Our analyses were also limited to trials published or reported in English, and it is therefore possible that the inclusion of RCT results that are only available in other languages might have altered our conclusions. Finally, our results must also be viewed with sample size limitations in mind. Although the number of studies available for computing estimates of treatment effects and SESs was limited for some measures, all of the mean estimates we have presented are based on the results of at least five trials[25] and the treatment effect estimates represent total numbers of patients ranging from 1,296 (clinician treatment response) to 13,486 (WOMAC function subscale). Nevertheless, it is important to recognize that there were a substantial number of trials in the database that did not report information from which an appropriate measure of variability could be determined for calculating treatment effect and SES estimates. This made the sample sizes for our meta-analyses considerably smaller than if there had been more complete reporting of the results of the clinical trials in the database. Improvements in the quality of the reporting of clinical trial results are needed to facilitate meta-analyses such as those performed here.

The assay sensitivity of an outcome measure is a function of the separation between measured improvement in the active treatment group and in the placebo group. It is widely appreciated that substantial improvements in pain occur in the placebo groups of OA trials[20,49], and "excessive" placebo group improvement for an outcome measure could compromise its responsiveness. Benefit in placebo groups can be due to multiple factors alone and in combination, including placebo effects, natural history, regression to the mean, and various subject, study site, and research design factors. In future research, it would be worthwhile to examine whether different types of outcome measures vary in the extent to which improvement is demonstrated with placebo treatment. Identifying specific outcome measures that are less responsive to placebo treatment than are other measures (while showing comparable responsiveness as the other measures to active treatments) has the potential to show greater treatment effects and thereby improve the assay sensitivity of analgesic trials[19].

There is little question that our analyses will need to be updated in several years. One important reason for this is that new outcome measures and new approaches to evaluating outcome in RCTs of pain in OA are being developed. For example, recent research has examined electronic pain diaries of various types[50], including mobile phones[51], and such methods may increase the convenience of collecting more frequent, and therefore, more reliable pain ratings (although there is little evidence to date that these measures show greater responsiveness to treatment). There has also been increasing attention to evaluating the clinical importance of outcome measures[14,52], including the identification of low or acceptable levels of symptoms[53,54]. In addition, the importance of considering the patient's perspective has been emphasized[15,55], and initial attempts have been made to develop patient-centered outcome measures that assess the specific treatment goals of individual patients[56–58]. These efforts may lead to the identification of outcome measures with greater reliability, validity, and responsiveness, which could increase the assay sensitivity of clinical trials of treatments for OA.

Declaration of funding and role of funding source

The OARSI FDA OA Initiative received financial support from the following professional organization:

American College of Rheumatology.

Additionally the OARSI FDA OA Initiative received financial support from the following companies: Amgen, ArthroLab, AstraZeneca, Bayer Healthcare, Chondrometrics, CombinatoRx, Cypress, BioScience, DePuy Mitek, Expanscience, 4QImaging, Genevrier/IBSA, Genzyme, King (Alpharma), Merck, Merck Serono, NicOx, Pfizer, Rottapharm, Smith & Nephew, Wyeth.

While individuals from pharmaceutical, biotechnology and device companies actively participated in on-going working group discussions, due to the conflict of interest policy enacted by OARSI, these individuals were not allowed to vote on the final recommendations made by OARSI to the Food and Drug Administration.

Author contributions

Conception and design of the study: RHD, SP, DCT, MPM, AG, LSS, JTF, NPK.

Acquisition and analysis of data: RHD, SP, MPM.

Interpretation of data: RHD, SP, DCT, MPM, AG, LSS, JTF, NPK.

Drafting the manuscript: RHD, SP, DCT.

Revision of manuscript for important content: RHD, SP, DCT, MPM, AG, LSS, JTF, NPK.

Final approval of the submitted manuscript: RHD, SP, DCT, MPM, AG, LSS, JTF, NPK.

Conflicts of interest

The authors do not have financial conflicts of interest related to the material presented in this article.

Appendix I

Treatment vs. placebo group differences, SESs, and study outcomes for osteoarthritis outcome measures for all treatments and arms

	Pain		Pain (with activity)		WOMAC pain subscale		WOMAC pain walking		≥30% pain reduction (% patients)	
	N*	Mean†	N	Mean	N	Mean	N	Mean	N	Mean
Treatment vs placebo group difference, unweighted	83	10.9	30	10.7	99	8.6	21	9.8	14	14.8
SES, unweighted	22	0.32	8	0.25	48	0.42	9	0.44		
% positive‡	101	75.3	38	86.8	118	76.3	21	90.5	15	66.7

	Patient global		Patient treatment response		Patient responder (% patients)		Clinician global		Clinician treatment response		Clinician responder (% patients)	
	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
Treatment vs placebo group difference, unweighted	66	11.5	20	15.5	61	23.7	50	10.6	11	18.4	35	24.3
SES, unweighted	34	0.39	9	0.66			29	0.37	5	0.74		
% positive‡	93	74.1	26	80.1	63	84.1	68	80.9	14	71.4	34	88.2

	WOMAC function subscale		Lequesne algofunctional index		WOMAC stiffness subscale		WOMAC total score		OMERACT–OARSI responder criteria (% patients)	
	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
Treatment vs placebo group difference, unweighted	93	7.6	32	5.8	82	7.8	65	7.8	18	13.2
SES, unweighted	47	0.35	10	0.39	46	0.33	29	0.34		
% positive‡	120	70.0	40	62.5	104	65.4	78	71.8	23	82.6

* Number of active treatment arms (one value per treatment).

† Confidence intervals are not provided because their calculation would assume the statistical independence of the results from all of the trials, which does not hold in this case due to the inclusion of multiple treatment comparisons with the same placebo group in some of the trials.

‡ For "% positive," the outcome for each treatment arm was categorized as positive if the comparison with placebo for a given measure yielded a statistically significant ($p \leq 0.05$) treatment effect, and the percentage of positive outcomes was calculated by dividing the number of treatment arms (one for each active treatment per trial) with a positive outcome by the total number of treatment arms. The outcome for multiple arms of the same treatment in a trial was categorized as positive if at least one of the arms was positive. The number of treatment arms for this variable can be higher than for the unweighted treatment effect because some studies provided values for outcome but insufficient information to compute treatment effects.

References

1. Bjordal JM, Klovning A, Ljunggren AE, Slørdal L. Short-term efficacy of pharmacotherapeutic interventions in osteoarthritic knee pain: a meta-analysis of randomised placebo-controlled trials. Eur J Pain 2007;11:125–38.
2. Zhang W, Nuki G, Moskowitz RW, Abramson S, Altman RD, Arden NK, *et al.* OARSI recommendations for the management of hip and knee osteoarthritis, part III: changes in evidence following systematic cumulative update of research published through January 2009. Osteoarthritis Cartilage 2010;18:476–99.
3. U.S. Department of Health and Human Services. Guidance for industry: E10 choice of control group and related issues in clinical trials, 2001; Available at: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm073139.pdf.
4. U.S. Department of Health and Human Services. Guidance for industry: clinical development programs for drugs, devices, and biological products intended for the treatment of osteoarthritis (OA), 1999; Available at: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071579.pdf.
5. Pham T, van der Heijde D, Altman RD, Anderson JJ, Bellamy N, Hochberg M, *et al.* OMERACT–OARSI initiative: Osteoarthritis Research Society International set of responder criteria for osteoarthritis clinical trials revisited. Osteoarthritis Cartilage 2004;12:389–99.
6. European Medicines Agency. Guideline on clinical investigation of medicinal products used in the treatment of osteoarthritis, 2009; Available at: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003443.pdf.
7. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15:1833–40.
8. Lequesne MG, Mery C, Samson M, Gerard P. Indexes of severity for osteoarthritis of the hip and knee: validation: value in comparison with other assessment tests. Scand J Rheumatol 1987;65(Suppl):85–9.
9. Altman R, Brandt K, Hochberg M, Moskowitz R, Bellamy N, Bloch DA, *et al.* Design and conduct of clinical trials in patients with osteoarthritis: recommendations from a task force of the Osteoarthritis Research Society. Osteoarthritis Cartilage 1996;4:217–43.
10. Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P, *et al.* Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. J Rheumatol 1997;24:799–802.
11. Dougados M, Leclaire P, van der Heijde D, Bloch DA, Bellamy N, Altman RD. Response criteria for clinical trials on osteoarthritis of the knee and hip: a report of the Osteoarthritis Research Society International Standing Committee for Clinical Trials Response Criteria Initiative. Osteoarthritis Cartilage 2000;8:395–403.
12. Maheu E, Altman RD, Bloch DA, Doherty M, Hochberg M, Mannoni A, *et al.* Design and conduct of clinical trials in patients with osteoarthritis of the hand: recommendations from a task force of the Osteoarthritis Research Society International. Osteoarthritis Cartilage 2006;14:303–22.
13. Turk DC, Dworkin RH, Allen RR, Bellamy N, Brandenburg N, Carr DB, *et al.* Core outcome domains for chronic pain

clinical trials: IMMPACT recommendations. Pain 2003;106: 337–45.
14. Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, *et al*. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 2005;113:9–19.
15. Turk DC, Dworkin RH, Burke LB, Gershon R, Rothman M, Scott J, *et al*. Developing patient-reported outcome measures for pain clinical trials: IMMPACT recommendations. Pain 2006;125:208–15.
16. Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, *et al*. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J Pain 2008;9:105–21.
17. Dworkin RH, Turk DC, McDermott MP, Peirce-Sandner S, Burke LB, Cowan P, *et al*. Interpreting the clinical importance of group differences in chronic pain clinical trials: IMMPACT recommendations. Pain 2009;146:238–44.
18. Dworkin RH, Turk DC, Peirce-Sandner S, McDermott MP, Farrar JT, Hertz S, *et al*. Placebo and treatment group responses in postherpetic neuralgia vs. painful diabetic peripheral neuropathy clinical trials in the REPORT database. Pain 2010;150:12–6.
19. Dworkin RH, Turk DC, Katz NP, Rowbotham MC, Peirce-Sandner S, Cerny I. Evidence-based clinical trial design for chronic pain pharmacotherapy: a blueprint for ACTION. Pain 2011;152(Suppl):S107–15.
20. Zhang W, Robertson J, Jones AC, Dieppe PA, Doherty M. The placebo effect and its determinants in osteoarthritis: meta-analysis of randomised controlled trials. Ann Rheum Dis 2008;67:1716–23.
21. Trijau S, Avouac J, Escalas C, Gossec L, Dougados M. Influence of flare design on symptomatic efficacy of non-steroidal anti-inflammatory drugs in osteoarthritis: a meta-analysis of randomized placebo-controlled trials. Osteoarthritis Cartilage 2010;18:1012–8.
22. Cochrane Handbook for Systematic Reviews of Interventions, Version 5.0.2, 2009; Available at: http://www.mrc-bsu.cam.ac.uk/cochrane/handbook.
23. National Collaborating Centre for Chronic Conditions. Osteoarthritis: National Clinical Guideline for Care and Management in Adults. London: Royal College of Physicians; 2008.
24. Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, *et al*. OARSI recommendations for the management of hip and knee osteoarthritis, part II: OARSI evidence-based, expert consensus guidelines. Osteoarthritis Cartilage 2008;16:137–62.
25. Wood L, Egger M, Gluud LL, Schulz KF, Jüni P, Altman DG, *et al*. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ 2008;336:601–5.
26. Faucher M, Poiraudeau S, Lefevre-Colau MM, Rannou F, Fermanian J, Revel M. Algo-functional assessment of knee osteoarthritis: comparison of the test-retest reliability and construct validity of the WOMAC and Lequesne indexes. Osteoarthritis Cartilage 2002;10:602–10.
27. Normand SL. Meta-analysis: formulating, evaluating, combining, and reporting. Stat Med 1999;18:321–59.
28. Whitehead A. Meta-analysis of Controlled Clinical Trials. Chichester, UK: John Wiley; 2002.
29. Gentelle-Bonnassies S, Le Claire P, Mezieres M, Ayral X, Dougados M. Comparison of the responsiveness of symptomatic outcome measures in knee osteoarthritis. Arthritis Care Res 2000;13:280–5.
30. Krebs EE, Bair MJ, Damush TM, Tu W, Wu J, Kroenke K. Comparative responsiveness of pain outcome measures among primary care patients with musculoskeletal pain. Med Care 2010;48:1007–14.
31. Bellamy N. WOMAC: a 20-year experiential review of a patient-centered self-reported health status questionnaire. J Rheumatol 2002;29:2473–6.
32. Villanueva I, del Mar Guzman M, Javier Toyos F, Ariza-Ariza R, Navarro F. Relative efficiency and validity properties of a visual analogue vs a categorical scaled version of the Western Ontario and McMaster Universities Osteoarthritis (WOMAC) Index: Spanish versions. Osteoarthritis Cartilage 2004;12:225–31.
33. Bellamy N, Campbell J, Syrotuik J. Comparative study of self-rating pain scales in osteoarthritis patients. Curr Med Res Opin 1999;15:113–9.
34. Bolognese JA, Schnitzer TJ, Ehrich EW. Response relationship of VAS and Likert scales in osteoarthritis efficacy measurement. Osteoarthritis Cartilage 2003;11:499–507.
35. Lund I, Lundeberg T, Sandberg L, Budh CN, Kowalski J, Svensson E. Lack of interchangeability between visual analogue and verbal rating pain scales: a cross sectional description of pain etiology groups. BMC Med Res Methodol 2005;5:31.
36. Dworkin RH, Jensen MP, Gould E, Jones BA, Xiang Q, Galer BS, *et al*. Treatment satisfaction in osteoarthritis and low back pain: the role of pain, physical and emotional functioning, sleep, and adverse events. J Pain 2011;12:416–24.
37. Bingham III CO, Bird SR, Smugar SS, Xu X, Tershakovec AM. Responder analysis and correlation of outcome measures: pooled results from two identical studies comparing etoricoxib, celecoxib, and placebo in osteoarthritis. Osteoarthritis Cartilage 2008;16:1289–93.
38. Carville SF, Choy EHS. Systematic review of discriminating power of outcome measures used in clinical trials of fibromyalgia. J Rheumatol 2008;35:2094–105.
39. Dunkl PR, Taylor AG, McConnell GG, Alfano AP, Conaway MR. Responsiveness of fibromyalgia clinical trial outcome measures. J Rheumatol 2000;27:2683–91.
40. Angst F, Aeschlimann A, Steiner W, Stucki G. Responsiveness of the WOMAC osteoarthritis index as compared with the SF-36 in patients with osteoarthritis of the legs undergoing a comprehensive rehabilitation intervention. Ann Rheum Dis 2001;60:834–40.
41. Soohoo NF, Vyas RM, Samimi DB, Molina R, Lieberman JR. Comparison of the responsiveness of the SF-36 and WOMAC in patients undergoing total hip arthroplasty. J Arthroplasty 2007;22:1168–73.
42. Quessy SN, Rowbotham MC. Placebo response in neuropathic pain trials. Pain 2008;138:479–83.
43. Katz N. Methodological issues in clinical trials of opioids for chronic pain. Neurology 2005;65:S32–49.
44. Katz J, Finnerup NB, Dworkin RH. Clinical trial outcome in neuropathic pain: relationship to study characteristics. Neurology 2008;70:263–72.
45. Wandel S, Jüni P, Tendal B, Nüesch E, Villiger PM, Welton NJ, *et al*. Effects of glucosamine, chondroitin, or placebo in patients with osteoarthritis of hip or knee: network meta-analysis. BMJ 2010;341. c4675.
46. Pelletier J-P, Hochberg MC, du Souich P, Kahan A, Michel BA. Effect size is encouraging. BMJ 2010;341. c6328.
47. Rowbotham MC. The impact of selective publication on clinical research in pain. Pain 2008;140:401–4.
48. Rowbotham MC. The case for publishing 'negative' trials. Pain 2009;146:225–6.

49. Scott-Lennox JA, McLaughlin-Miley C, Lennox RD, Bohlig AM, Cutler BL, Yan C, *et al*. Stratification of flare intensity identifies placebo responders in a treatment efficacy trial of patients with osteoarthritis. Arthritis Rheum 2001;44:1599–607.
50. Allen KD, Coffman CJ, Golightly YM, Stechuchak KM, Voils CI, Keefe FJ. Comparison of pain measures among patients with osteoarthritis. J Pain 2010;11:522–7.
51. Bellamy N, Wilson C, Hendrikz J, Whitehouse SL, Patel B, Dennison S. Osteoarthritis Index delivered by mobile phone (m-WOMAC) is valid, reliable, and responsive. J Clin Epidemiol 2011;64:182–90.
52. Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N, *et al*. Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement. Ann Rheum Dis 2005;64:29–33.
53. Dougados M, Moore A, Yu S, Gitton X. Evaluation of the Patient Acceptable Symptom State in a pooled analysis of two multicentre, randomized, double-blind, placebo-controlled studies evaluating lumiracoxib and celecoxib in patients with osteoarthritis. Arthritis Res Ther 2007;9. R11.
54. Bellamy N, Bell MJ, Goldsmith CH, Lee S, Maschio M, Raynauld J-P, *et al*. BLISS index using WOMAC index detects between-group differences at low-intensity symptom states in osteoarthritis. J Clin Epidemiol 2010;63:566–74.
55. U.S. Department of Health and Human Services. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims, 2009; Available at: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf.
56. Ruta DA, Garratt AM, Russell IT. Patient centred assessment of quality of life for patients with four common conditions. Qual Health Care 1999;8:22–9.
57. Clinch J, Tugwell P, Wells G, Shea B. Individualized functional priority approach to the assessment of health related quality of life in rheumatology. J Rheumatol 2001;28:445–51.
58. Jolles BM, Buchbinder R, Beaton DE. A study compared nine patient-specific indices for musculoskeletal disorders. J Clin Epidemiol 2005;58:791–801.